

### IN THE SPECIFICATION

Please amend the written specification of the patent application as follows wherein newly added text is indicated with underlining and deleted text is marked with ~~strikethrough~~ or enclosed within [[double brackets]]:

Please amend the paragraph in written specification from line 9 to line 16 on page 5 as follows:

**Figure 1** illustrates the interconnections of the various components that may be used to deliver a composite video signal to individual viewers. Video sources **100** and **126** send video signals **102** and **128** ~~[[126]]~~ through a distribution network **104** to viewer's locations **111**. Additionally, multiple interactive video servers **106** and **116** send video, HTML, and other attachments **108**. The multiple feeds **110** are sent to several set top boxes **112**, **118**, and **122** connected to televisions **114**, **120**, and **124**, respectively. The set top boxes **112** and **118** may be interactive set top boxes and set top box **122** may not have interactive features.

Please amend the paragraph in written specification from line 22 to line 28 on page 5 as follows:

The interactive set top boxes **112** and **118** may communicate to the interactive video servers **106** and **116** ~~[[108]]~~ though the video distribution network **104** if the video distribution network supports two-way communication, such as with cable modems. Additionally, communication may be through other upstream communication networks **130**. Such upstream networks may include a dial up modem, direct Internet connection, or other communication network that allows communication separate from the video distribution network **104**.

Please amend the paragraph in written specification from line 29 of page 5 to line 6 on page 6 as follows:

Although **Figure 1** illustrates the use of interactive set-top boxes **112** and **118**, the present invention can be implemented without an interactive connection with an interactive video server, such as interactive video servers **106** and **116**. In that case, separate multiple video sources **100** and **126** and can provide multiple video feeds **110** to non-interactive set-top box **122** at the viewer's locations **111**. The difference between the interactive set top boxes **112** and **118** and the non-interactive set top box **122** is that the interactive set top boxes **112** and **118** incorporate the functionality to receive, format, and display interactive content and send interactive requests to the interactive video servers **106** and **116**.

Please amend the paragraph in written specification from line 7 to line 13 on page 6 as follows:

The set top boxes **112**, **118**, and **122** may receive and decode two or more video feeds and combine the feeds to produce a composite video signal that is displayed for the viewer. Such a composite video signal may be different for each viewer, since the video signals may be combined in several different manners. The manner in which the signals are combined is described in a “presentation description” ~~the presentation description~~. The presentation description may be provided through the interactive video servers **106** and **116** or through another server **132**. Server **132** may be a web server or a specialized data server.

Please amend the paragraph in written specification from line 25 of page 6 to line 17 on page 6 as follows:

The manner in which the video signals are to be combined is defined in the presentation description. The presentation description may be a separate file provided by the server **132**, the interactive video servers **106** and **116**, or may be embedded into one or more of the multiple feeds **110**. A plurality of presentation descriptions may be transmitted and program code

operating in a set top box may select one or more of the presentation descriptions based upon an identifier in the presentation description(s). This allows presentation descriptions to be selected that correspond to set top box requirements and/or viewer preferences or other information. Further, demographic information may be employed by upstream equipment to determine a presentation description version for a specific set top box or group of set top boxes and an identifier of the presentation description version(s) may then be sent to the set top box or boxes. Presentation descriptions may also be accessed across a network, such as the Internet, that may employ upstream communication on a cable system or other networks. In a similar manner, a set top box may access a presentation description across a network that corresponds to set top box requirements and/or viewer preferences or other information. And in a similar manner as described above, demographic information may be employed by upstream equipment to determine a presentation description version for a specific set top box or group of set top boxes and an identifier of the presentation description version(s) may then be sent to the set top box or boxes. The identifier may comprise a URL, filename, extension or other information that identifies the presentation description. Further, a plurality of presentation descriptions may be transferred to a set top box and a viewer may select versions of the presentation description. Alternatively, software program operating in the set top box may generate the presentation description and such generation may also employ viewer preferences or demographic information.

Please amend the paragraph in written specification from line 3 to line 16 on page 10 as follows:

The presentation description information **216** is the information necessary for the video combiner **232** to combine the various portions of multiple video signals to form a composite video image. The presentation description information **216** can take many forms[[,]] such as an ATVEF trigger, [[or]] a markup language description using HTML<sub>a</sub> or a similar format. Such information may be transmitted in a vertical blanking encoded signal that includes instructions as to the manner in which to combine the various video signals. For example, the presentation description may be encoded in the vertical blanking interval (VBI) of stream **206** **[[210]]**. The

presentation description may also include Internet addresses for connecting to enhanced video web sites. The presentation description information **216** may include specialized commands applicable to specialized set top boxes, or may contain generic commands that are applicable to a wide range of set top boxes. References made herein to the ATVEF specification are made for illustrative purposes only, and such references should not be construed as an endorsement, in any manner, of the ATVEF specification.

Please amend the paragraph in written specification from line 20 of page 16 to line 24 on page 17 as follows:

**Figure 5** depicts another set top box embodiment of the present invention. Set top box **500** comprises tuner/decoder **502**, decoder **504**, memory **506**, processor **508**, optional network interface **510**, video output unit **512**, and user interface **514**. Tuner/decoder **502** receives a broadcast that comprises at least two video signals. In one embodiment of **figure 5**, tuner/decoder **502** is capable of tuning at least two independent frequencies. In another embodiment of **figure 5**, tuner/decoder **502** decodes at least two video signals contained within a broadcast band, as may occur with QAM or QPSK transmission over analog television channel bands or satellite bands. "Tuning" of video signals may comprise identifying packets with predetermined PID (packet identifier Identifiers) values or a range thereof and forwarding such packets to processor **508** or to decoder **504**. For example, data packets may be transferred to decoder **504** and control packets may be transferred to processor **508**. Data packets may be discerned from control packets through secondary PIDs or through PID values in a predetermined range. Decoder **504** processes packets received from tuner/decoder **502** and generates and stores image and/or audio information in memory **506**. Image and audio information may comprise various information types common to DCT based image compression methods, such as MPEG and motion JPEG, for example, or common to other compression methods such as wavelets and the like. Audio information may conform to MPEG or other formats such as those developed by Dolby Laboratories and THX as are common to theaters and home entertainment systems. Decoder **504** may comprise one or more decoder chips to provide sufficient processing capability to process two or more video streams substantially

simultaneously. Control packets provided to processor **508** may include presentation description information. Presentation description information may also be accessed employing network interface **510**. Network interface **510** may comprise any type of network that provides access to a presentation description including modems, cable modems, DSL modems, upstream channels in a set top box and the like. Network interface **510** may also be employed to provide user responses to interactive content to [[a]] an associated server or other equipment. Processor **508** employs the presentation description to control combination of the image and/or audio information stored in memory **506**. Combination may employ processor **508**, decoder **504**, or a combination of processor **508** and decoder **504**. Combined image and or audio information, as created employing the presentation description, is supplied to video output unit **512** that produces and output signal for a television, monitor, or other type of display. The output signal may comprise composite video, S-video, RGB, or any other format. User interface **514** supports a remote control, mouse, keyboard, or other input device. User input may serve to select versions of a presentation description or to modify a presentation description.

Please amend the paragraph in written specification from line 25 of page 17 to line 11 on page 18 as follows:

**Figure 6** depicts a sequence of steps **600** employed to create a combined image at a user's set top box. At step **602** a plurality of video signals are received. These signals may contain digitally encoded image and audio data. At step **604** a presentation description is accessed. The presentation description may be part of a broadcast signal, or may be accessed across a network. At step **606**, at least two of the video signals are decoded and image data and audio data (if present) for each video signal is stored in a memory of the set top box. At step **608**, portions of the video images and optionally portions of the audio data are combined in accordance with the presentation description. The combination of video images and optionally audio data may produce combined data in the memory of [[f]] the set top box, or such combination may be performed "on the fly" wherein real-time combination is performed and the output provided to step **610**. For example, if a mask is employed to select between portions of two images, non-sequential addressing of the set top box memory may be employed to access portions of each

image in a real-time manner, eliminating the need to create a final display image in set top box memory. At step 610 the combined image and optionally combined audio are output to a presentation device such as a television, monitor, or other display device. Audio may be provided to the presentation device or to an amplifier, stereo system, or other audio equipment.